Fig. 1

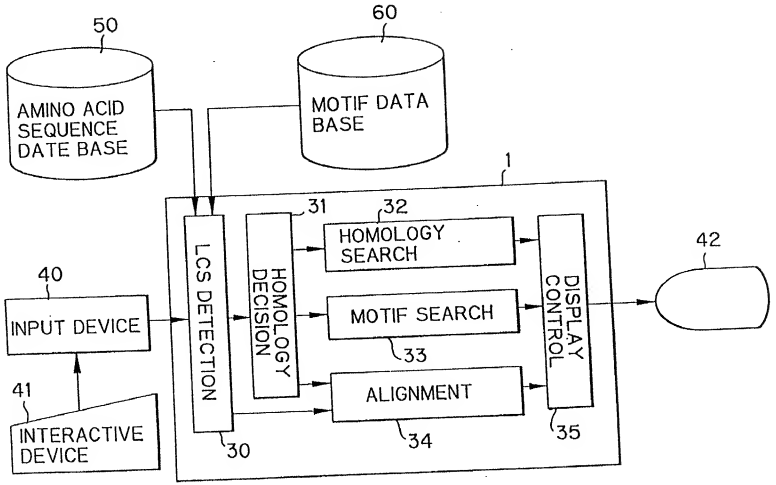

50
AMINO ACID SEQUENCE DATE BASE
60
MOTIF DATA BASE
1
31
32
HOMOLOGY SEARCH
LCS DETECTION
HOMOLOGY DECISION
MOTIF SEARCH
33
ALIGNMENT
34
DISPLAY CONTROL
35
42
40
INPUT DEVICE
41
INTERACTIVE DEVICE
30

Fig. 2

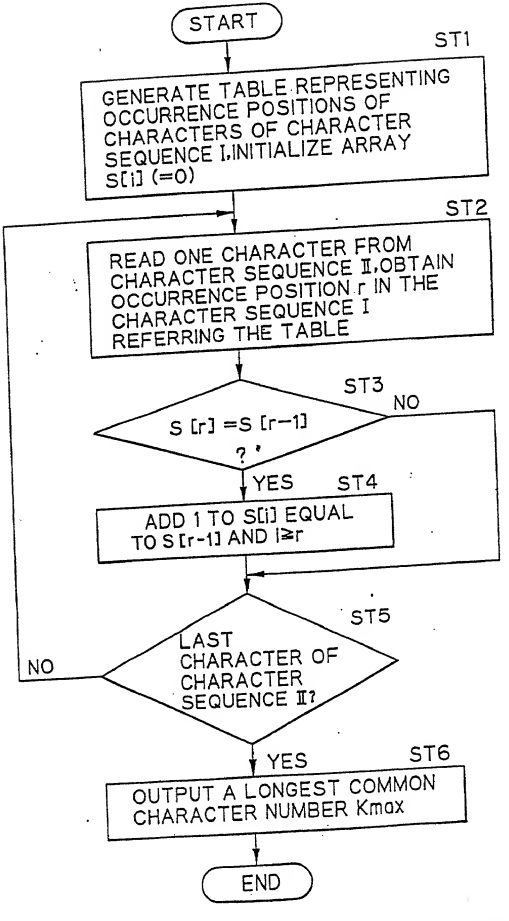

START
ST1
GENERATE TABLE REPRESENTING OCCURRENCE POSITIONS OF CHARACTERS OF CHARACTER SEQUENCE I,INITIALIZE ARRAY S[i] (=0)
ST2
READ ONE CHARACTER FROM CHARACTER SEQUENCE II,OBTAIN OCCURRENCE POSITION r IN THE CHARACTER SEQUENCE I REFERRING THE TABLE
ST3
S [r] =S [r−1] ?
NO
YES
ST4
ADD 1 TO S[i] EQUAL TO S [r-1] AND i≧r
ST5
LAST CHARACTER OF CHARACTER SEQUENCE II?
NO
YES
ST6
OUTPUT A LONGEST COMMON CHARACTER NUMBER Kmax
END

Fig. 3

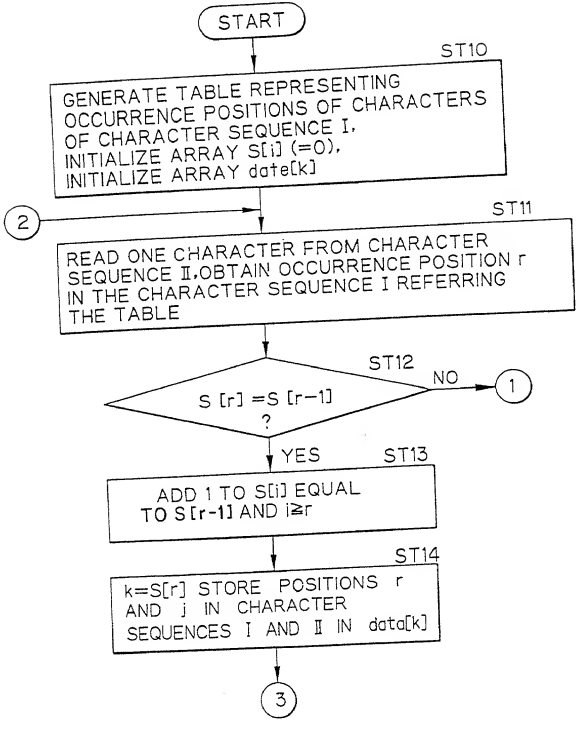

START
ST10
GENERATE TABLE REPRESENTING OCCURRENCE POSITIONS OF CHARACTERS OF CHARACTER SEQUENCE I,
INITIALIZE ARRAY S[i] (=0),
INITIALIZE ARRAY date[k]
2
ST11
READ ONE CHARACTER FROM CHARACTER SEQUENCE II, OBTAIN OCCURRENCE POSITION r IN THE CHARACTER SEQUENCE I REFERRING THE TABLE
ST12
S [r] =S [r−1] ?
NO
1
YES
ST13
ADD 1 TO S[i] EQUAL TO S[r-1] AND i≧r
ST14
k=S[r] STORE POSITIONS r AND j IN CHARACTER SEQUENCES I AND II IN data[k]
3

Fig. 4

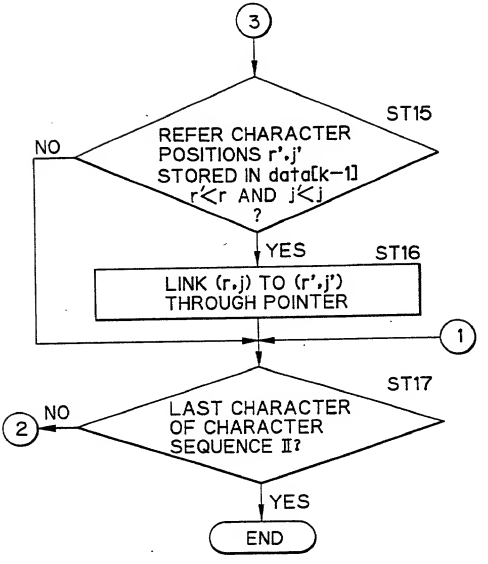
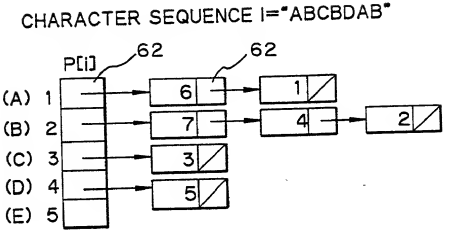

3
ST15
REFER CHARACTER POSITIONS r',j' STORED IN data[k−1] r'<r AND j'<j ?
NO
YES
ST16
LINK (r,j) TO (r',j') THROUGH POINTER
1
ST17
LAST CHARACTER OF CHARACTER SEQUENCE II?
2
NO
YES
END

Fig. 5



T0E220 '1505T660

Fig. 6

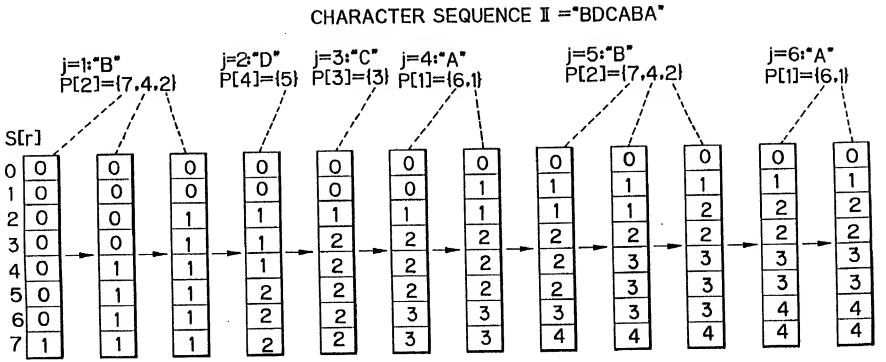


Fig. 7

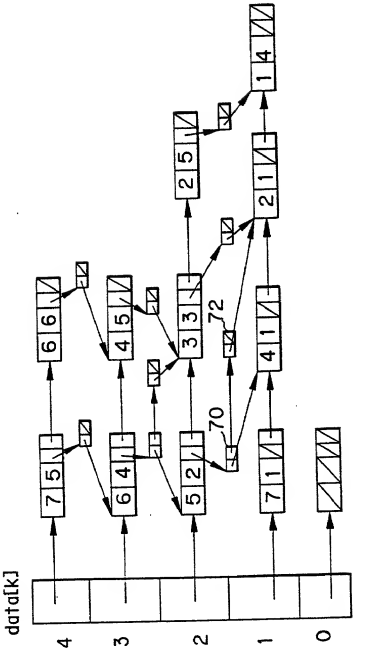


Fig. 8

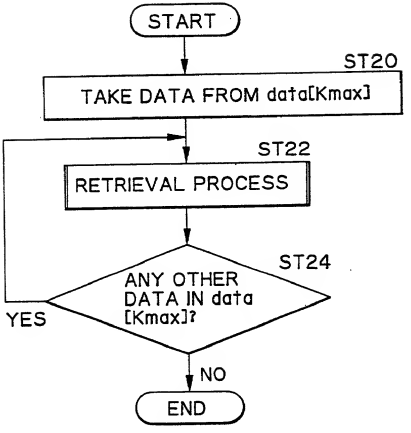
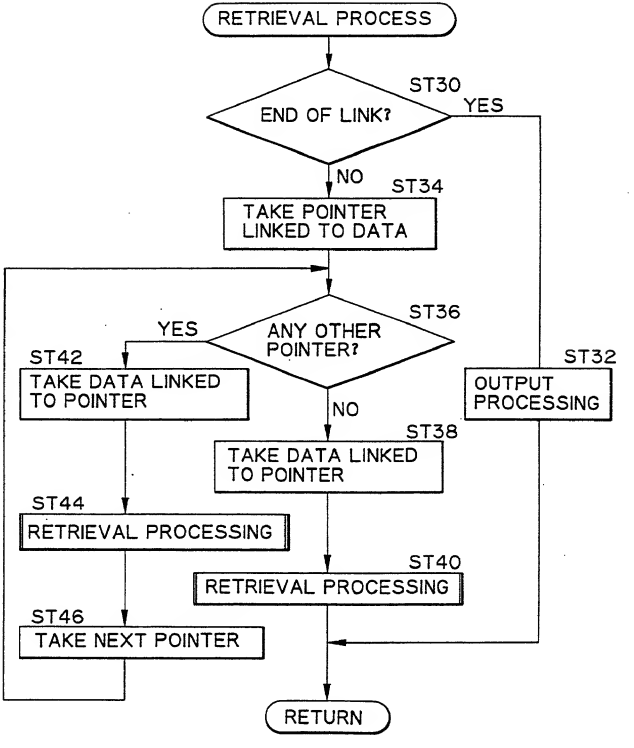

START
ST20
TAKE DATA FROM data[Kmax]
ST22
RETRIEVAL PROCESS
ST24
ANY OTHER DATA IN data [Kmax]?
YES
NO
END

Fig. 9


RETRIEVAL PROCESS
ST30
END OF LINK?
YES
NO
ST34
TAKE POINTER LINKED TO DATA
ST36
ANY OTHER POINTER?
YES
NO
ST42
TAKE DATA LINKED TO POINTER
ST32
OUTPUT PROCESSING
ST38
TAKE DATA LINKED TO POINTER
ST44
RETRIEVAL PROCESSING
ST40
RETRIEVAL PROCESSING
ST46
TAKE NEXT POINTER
RETURN

T0E220'1SC0T660

Fig. 10

human : GDVEKGKKIFIMKCSQCHTVEGGKHKTGPNLHGLFGRK
bacterium : EGDAAAGEKVSKKCLACHTFDQGGANKVGPNPNLFGVF

LCS : GD{x3.3}G{x0.1}K{x0.2}K{x4.0}KC{x2.2}CHT{x3.3}GG{x2.2}K
GD{x1.4}E{x0.2}K{x0.2}K{x0.4}KC{x2.2}CHT{x3.3}GG{x2.2}K

homology : 47%

WO 99/05284 PCT/US98/15660

Fig. 11

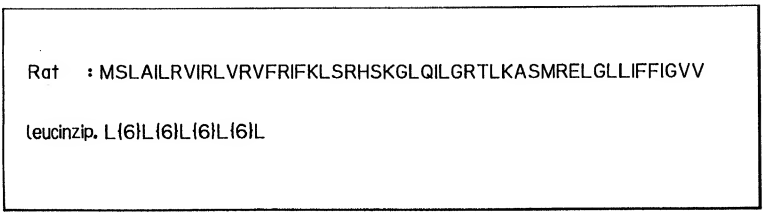


Fig. 12

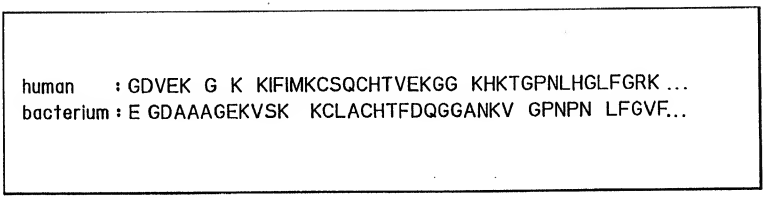
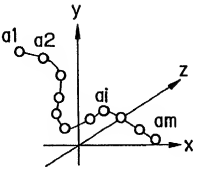
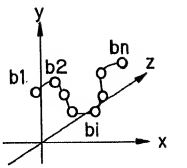

human : GDVEK G K KIFIMKCSQCHTVEKGG KHKTGPNLHGLFGRK ...
bacterium : E GDAAAGEKVSK KCLACHTFDQGGANKV GPNPN LFGVF...

Fig. 13 A



A={a1,a2,...ai,...am}

Fig. 13 B



B={b1,b2,...bj,...bn}

Fig. 13C

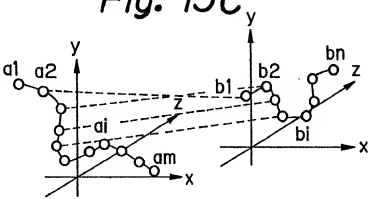


Fig. 13 D

$$r.m.s.d = \frac{\sqrt{\sum_{k=1}^{n} w_k (Ub_k - a_k)^2}}{n}$$

Fig. 15

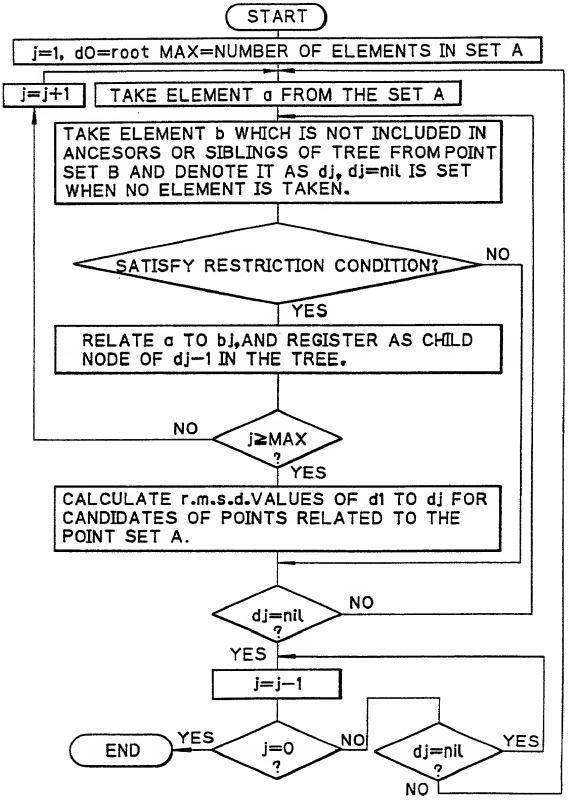

START
j=1, d0=root MAX=NUMBER OF ELEMENTS IN SET A
j=j+1
TAKE ELEMENT a FROM THE SET A
TAKE ELEMENT b WHICH IS NOT INCLUDED IN ANCESORS OR SIBLINGS OF TREE FROM POINT SET B AND DENOTE IT AS dj, dj=nil IS SET WHEN NO ELEMENT IS TAKEN.
SATISFY RESTRICTION CONDITION?
NO
YES
RELATE a TO bj,AND REGISTER AS CHILD NODE OF dj-1 IN THE TREE.
NO
j≧MAX ?
YES
CALCULATE r.m.s.d.VALUES OF d1 TO dj FOR CANDIDATES OF POINTS RELATED TO THE POINT SET A.
dj=nil ?
NO
YES
j=j-1
END
YES
j=0 ?
NO
dj=nil ?
YES
NO

Fig. 14 A

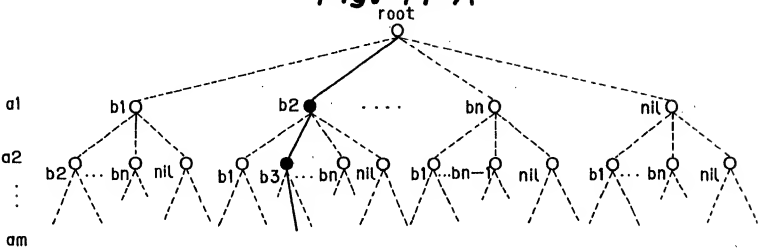


Fig. 14 B

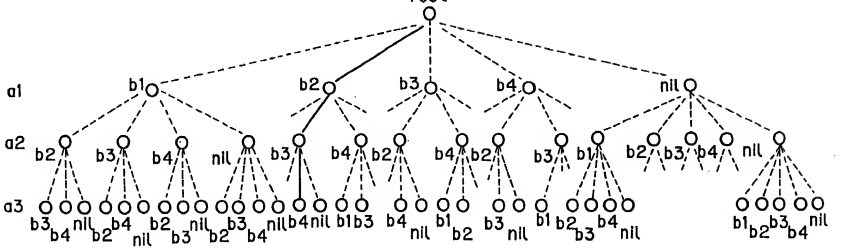


Fig. 16 A

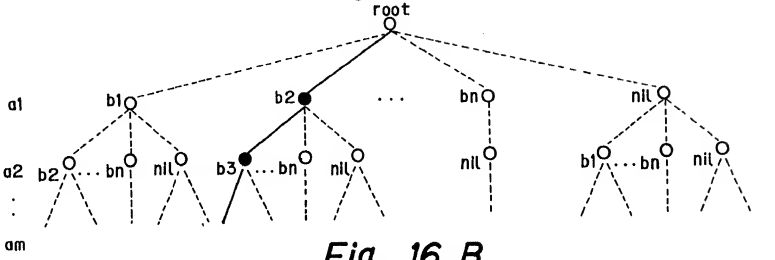


Fig. 16 B

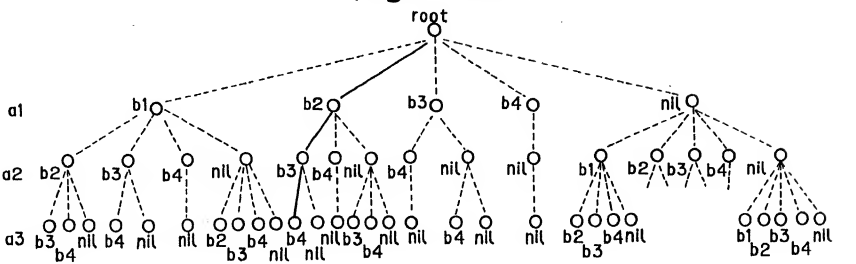


Fig. 17

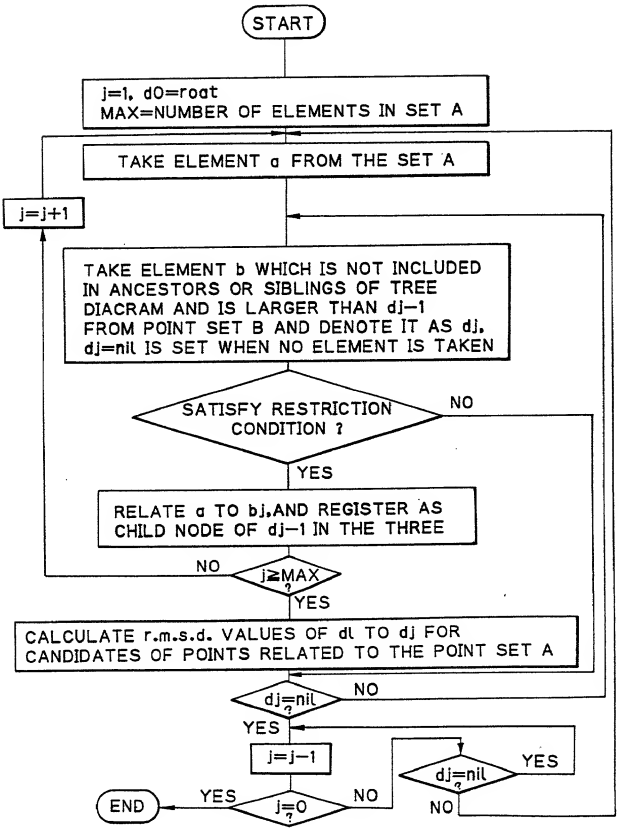

START
j=1, d0=root
MAX=NUMBER OF ELEMENTS IN SET A
TAKE ELEMENT a FROM THE SET A
j=j+1
TAKE ELEMENT b WHICH IS NOT INCLUDED IN ANCESTORS OR SIBLINGS OF TREE DIACRAM AND IS LARGER THAN dj−1 FROM POINT SET B AND DENOTE IT AS dj. dj=nil IS SET WHEN NO ELEMENT IS TAKEN
SATISFY RESTRICTION CONDITION ?
NO
YES
RELATE a TO bj, AND REGISTER AS CHILD NODE OF dj−1 IN THE THREE
NO
j≧MAX ?
YES
CALCULATE r.m.s.d. VALUES OF d1 TO dj FOR CANDIDATES OF POINTS RELATED TO THE POINT SET A
dj=nil ?
NO
YES
j=j−1
dj=nil ?
YES
NO
END
YES
j=0 ?
NO

Fig. 18

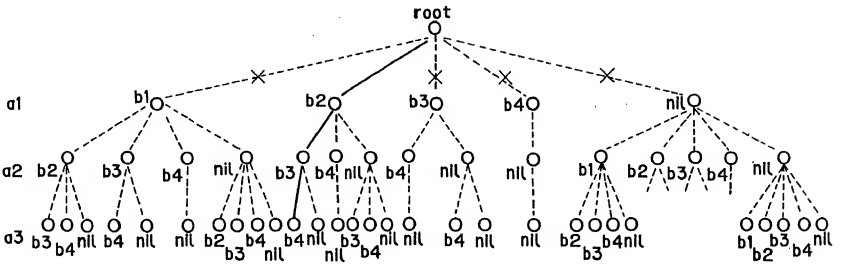


Fig. 19 A

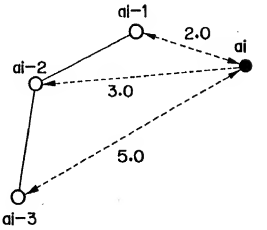


Fig. 19 B

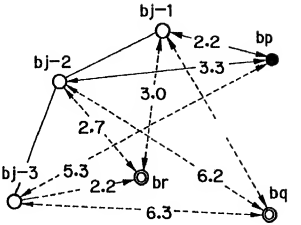


Fig. 20 A

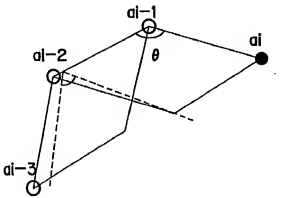


Fig. 20 B

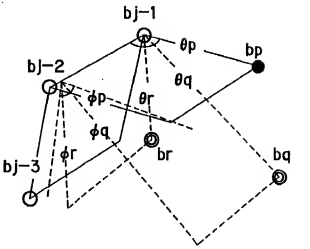
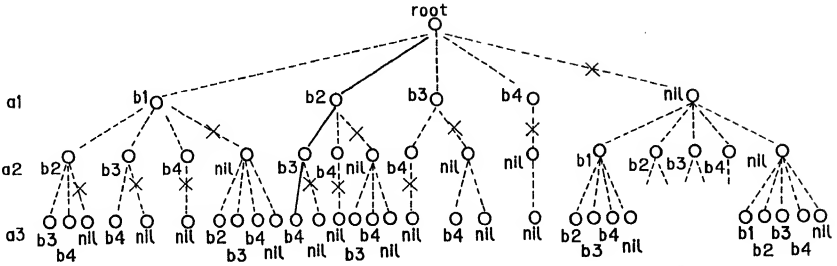


Fig. 21



09/10051,02/20/01

Fig. 22

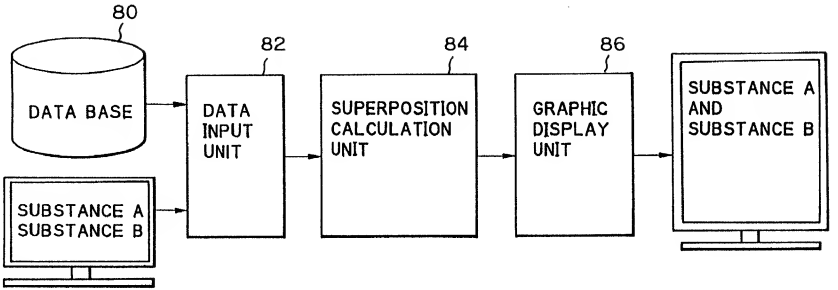


Fig. 23 A

1	T	E	E	Q	I	A	E	F	K	E	A	F	S	L	F	D
21	G	T	I	T	T	K	E	L	G	T	V	M	R	S	L	G
41	E	A	E	L	Q	D	M	I	N	E	V	D	A	D	G	N
61	F	P	E	F	L	T	M	M	A	R	K	M	K	D	T	D
81	I	R	E	A	F	R	V	F	D	K	D	G	N	G	Y	I
101	L	R	H	V	M	T	N	L	G	E	K	L	T	D	E	E
121	I	R	E	A	N	I	D	G	D	G	Q	V	N	Y	E	E
141	M	T	A													

1	A	F	S	L	F	D	K	D	G	D						
21	V	M	R	S	L	G	Q	N	P	T						
41	V	D	A	D	G	N	G	T	I	D						
61	K	M	K	D	T	D	S	E	E	E						
81	D	G	N	G	Y	I	S	A	A	E						
101	K	L	T	D	E	E	V	D	E	M						
121	Q	V	N	Y	E	E	F	V	Q	M						

AMINO ACID SEQUENCE OF CALMODULIN
(EXCERPT FROM PDB)

Fig. 23 B

1	A	M	D	Q	Q	A	E	A	R	A	F	L	S	E	E	M
21	K	A	A	F	D	M	F	D	A	D	G	G	G	D	I	S
41	G	T	V	M	R	M	L	G	Q	N	P	T	K	E	E	L
61	E	E	V	D	E	D	G	S	G	T	I	D	F	E	E	F
81	V	R	Q	M	K	E	D	A	K	G	K	S	E	E	E	L
101	R	I	F	D	K	N	A	D	G	F	I	D	I	E	E	L
121	R	A	T	G	E	H	V	T	E	E	D	I	E	D	L	M
141	K	N	N	D	G	R	I	D	F	D	E	F	L	K	M	M
161																

1	A	E	F	L	S	E	E	M	I	A	E	F				
21	K	K	S	E	E	E	L	A	D	C	F					
41	P	T	K	E	E	L	D	A	I	I						
61	I	D	F	E	E	F	L	V	M							
81	K	S	E	E	E	L	A	D	C	F						
101	I	D	I	E	E	L	G	E	I	L						
121	D	I	E	D	L	M	K	D	S	D						
141	E	F	L	K	M	M	E	G	V	Q						

AMINO ACID SEQUENCE OF TROPONIN C
(EXCERPT FROM PDB)

0992009A.072301

Fig. 24 A

CALMODULIN

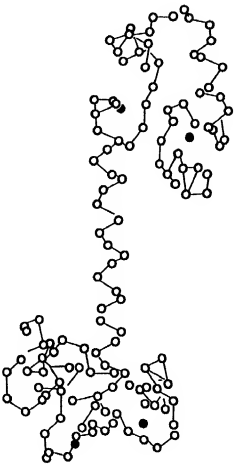
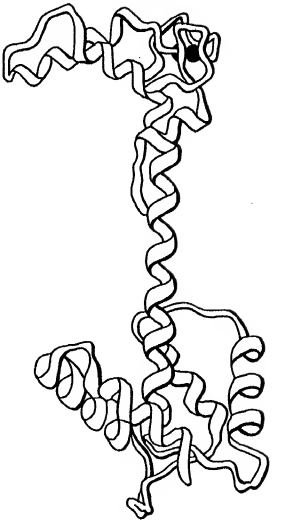


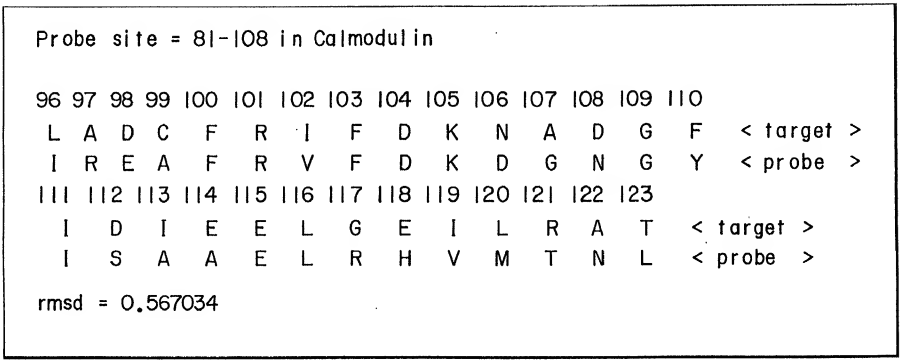
Fig. 24 B

TROPONIN C



T02270 1500660

Fig. 25



TOSZZ0'HSOOT600

Fig. 26

Probe site = 81-108 and 117-143 in Calmodulin

96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	
L	A	D	C	F	R	I	F	D	K	N	A	D	G	F	< target >
I	R	E	A	F	R	V	F	D	K	D	G	N	G	Y	< probe >
111	112	113	114	115	116	117	118	119	120	121	122	123			
I	D	I	E	E	L	G	E	I	L	R	A	T	< target >		
I	S	A	A	E	L	R	H	V	M	T	N	L	< probe >		
132	133	134	135	136	137	138	139	140	141	142	143	144	145		
I	E	D	L	M	K	D	S	D	K	N	N	D	G	< target >	
V	D	E	M	I	R	E	A	N	I	D	G	D	G	< probe >	
146	147	148	149	150	151	152	153	154	155	156	157	158			
R	I	D	F	D	E	F	L	K	M	M	E	G	< target >		
Q	V	N	Y	E	E	F	V	Q	M	M	T	A	< probe >		

rmsd = 0.823665

T0E220 '1500T660

Fig. 27

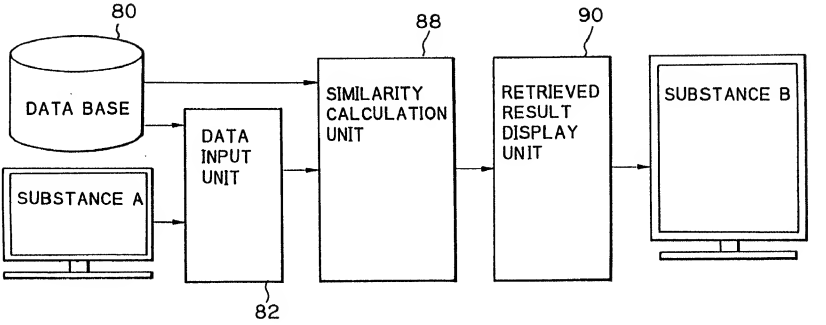
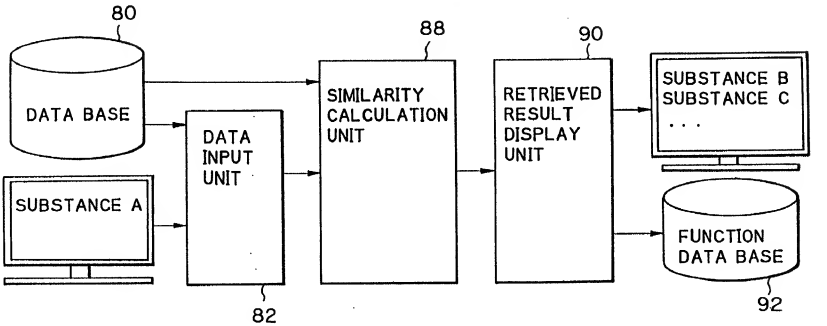


Fig. 28


80
DATA BASE
SUBSTANCE A
DATA INPUT UNIT
82
88
SIMILARITY CALCULATION UNIT
90
RETRIEVED RESULT DISPLAY UNIT
SUBSTANCE B
SUBSTANCE C
. . .
FUNCTION DATA BASE
92

EP 0 823 470 A1

Fig. 29

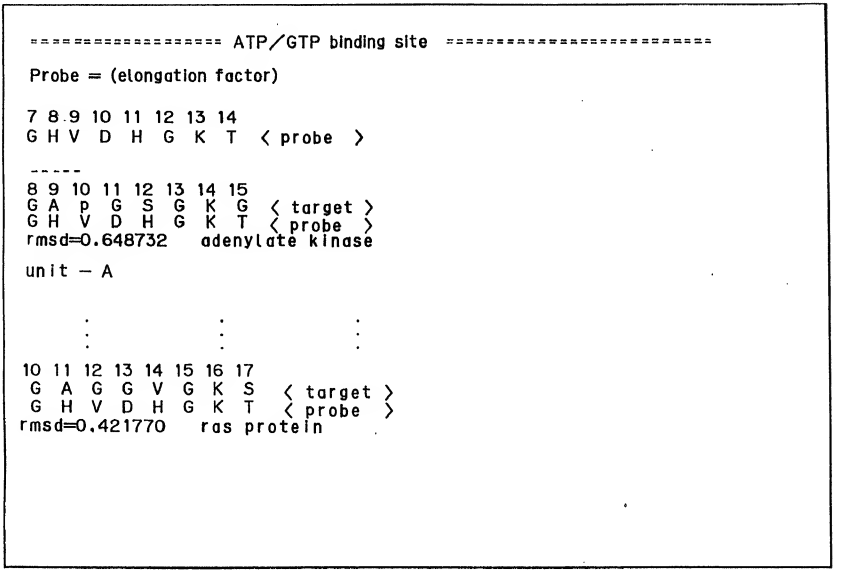


Fig. 30

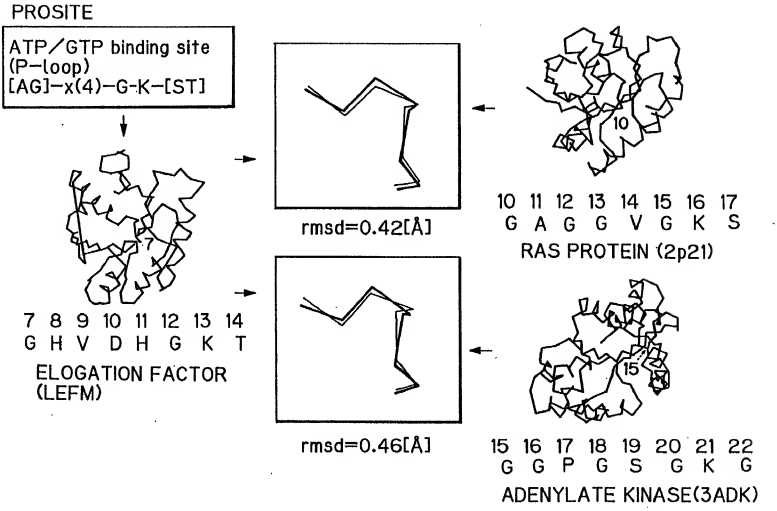

PROSITE
ATP/GTP binding site
(P-loop)
[AG]-x(4)-G-K-[ST]
7 8 9 10 11 12 13 14
G H V D H G K T
ELOGATION FACTOR
(LEFM)
rmsd=0.42[Å]
rmsd=0.46[Å]
10
10 11 12 13 14 15 16 17
G A G G V G K S
RAS PROTEIN (2p21)
15
15 16 17 18 19 20 21 22
G G P G S G K G
ADENYLATE KINASE(3ADK)

Fig. 31

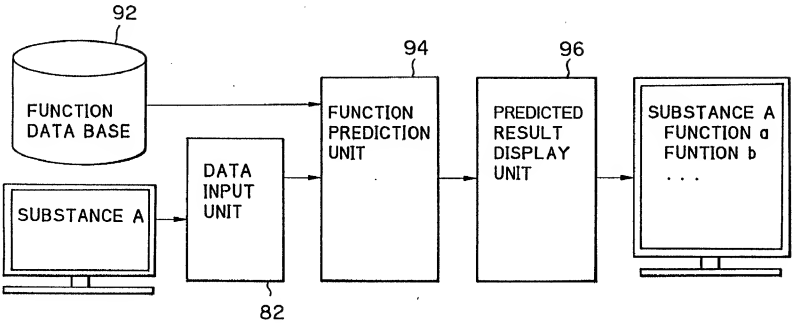
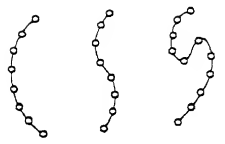
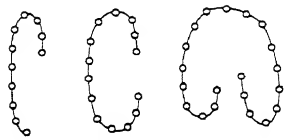

92
FUNCTION DATA BASE
SUBSTANCE A
DATA INPUT UNIT
82
94
FUNCTION PREDICTION UNIT
96
PREDICTED RESULT DISPLAY UNIT
SUBSTANCE A
FUNCTION a
FUNTION b
. . .

Fig. 32 A



LINEAR STRUCTURE

Fig. 32 B

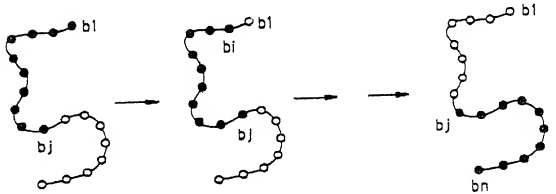


NON-LINEAR STRUCTURE

Fig. 33

WHEN f(x)=2x

A={a1.....am}



B={b1.....bi.....bj.....bn}

Fig. 34

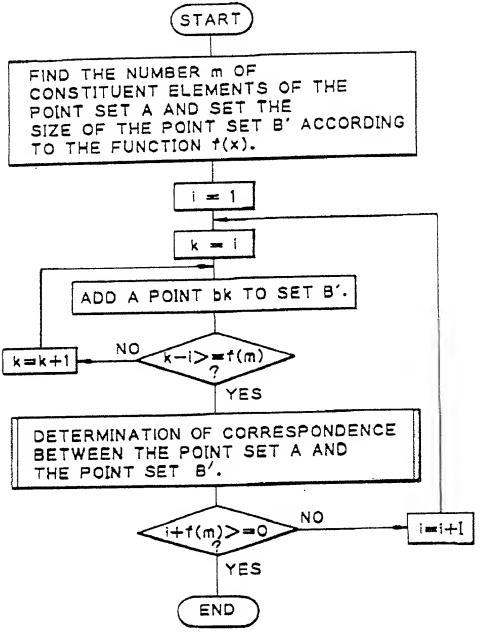
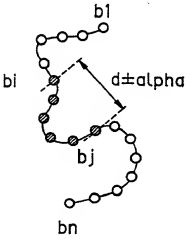

START
FIND THE NUMBER m OF CONSTITUENT ELEMENTS OF THE POINT SET A AND SET THE SIZE OF THE POINT SET B' ACCORDING TO THE FUNCTION f(x).
i = 1
k = i
ADD A POINT bk TO SET B'.
k-i>=f(m) ?
NO
k=k+1
YES
DETERMINATION OF CORRESPONDENCE BETWEEN THE POINT SET A AND THE POINT SET B'.
i+f(m)>=0 ?
NO
i=i+1
YES
END

Fig. 35 A

A={a1,a2,....,am}

Fig. 35 B



B={b1,....,bi,....,bj,....,bn}

Fig. 36

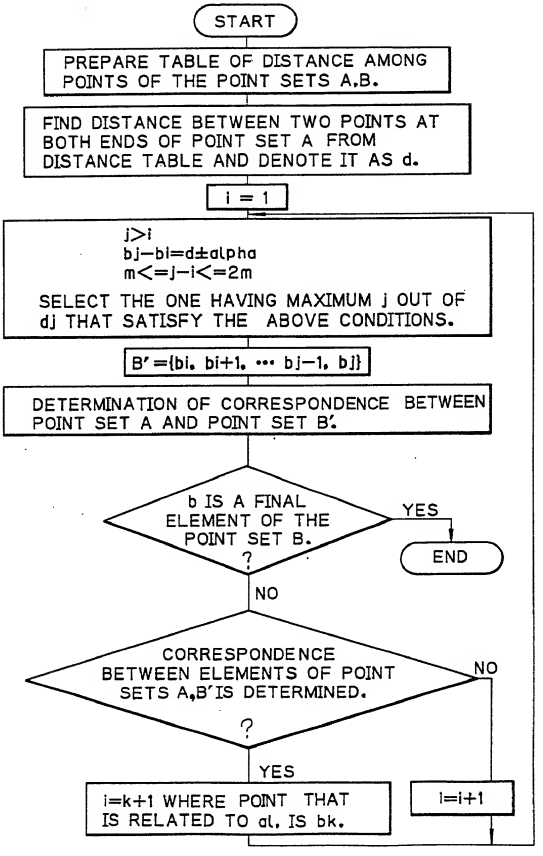

START
PREPARE TABLE OF DISTANCE AMONG POINTS OF THE POINT SETS A,B.
FIND DISTANCE BETWEEN TWO POINTS AT BOTH ENDS OF POINT SET A FROM DISTANCE TABLE AND DENOTE IT AS d.
i = 1
j>i
bj−bi=d±alpha
m<=j−i<=2m
SELECT THE ONE HAVING MAXIMUM j OUT OF dj THAT SATISFY THE ABOVE CONDITIONS.
B′={bi, bi+1, ••• bj−1, bj}
DETERMINATION OF CORRESPONDENCE BETWEEN POINT SET A AND POINT SET B′.
b IS A FINAL ELEMENT OF THE POINT SET B. ?
YES
END
NO
CORRESPONDENCE BETWEEN ELEMENTS OF POINT SETS A,B′IS DETERMINED. ?
NO
YES
i=k+1 WHERE POINT THAT IS RELATED TO ai, IS bk.
i=i+1

Fig. 37

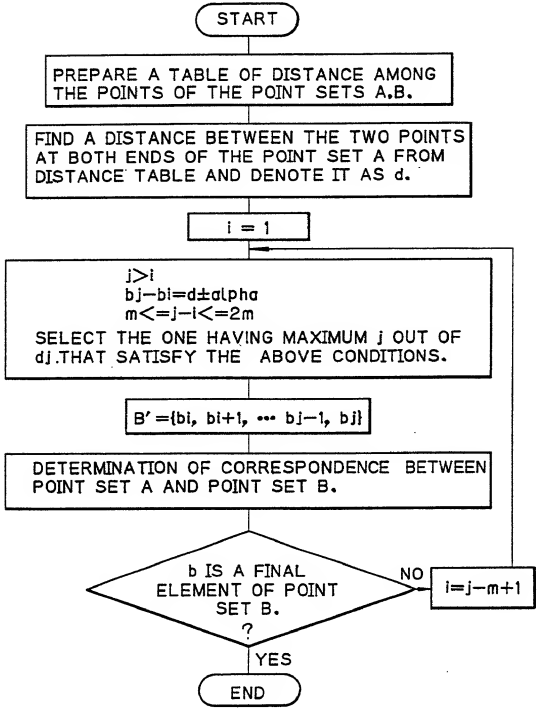

START
PREPARE A TABLE OF DISTANCE AMONG THE POINTS OF THE POINT SETS A,B.
FIND A DISTANCE BETWEEN THE TWO POINTS AT BOTH ENDS OF THE POINT SET A FROM DISTANCE TABLE AND DENOTE IT AS d.
i = 1
j>i
bj−bi=d±alpha
m<=j−i<=2m
SELECT THE ONE HAVING MAXIMUM j OUT OF dj.THAT SATISFY THE ABOVE CONDITIONS.
B' ={bi, bi+1, ••• bj−1, bj}
DETERMINATION OF CORRESPONDENCE BETWEEN POINT SET A AND POINT SET B.
b IS A FINAL ELEMENT OF POINT SET B. ?
NO
i=j−m+1
YES
END

Fig. 38 A

1	I	V	G	G	Y	T	C	C	A	N	T	V	P	Y	Q	V	S	L	N	S
21	G	Y	H	F	C	G	G	S	L	I	N	S	Q	W	V	V	S	A	A	H
41	C	Y	K	S	G	I	Q	V	R	L	G	E	D	N	I	N	V	V	E	G
61	N	E	Q	F	I	S	A	S	K	S	I	V	H	P	S	Y	N	S	N	T
81	L	N	N	D	I	M	L	I	K	L	K	S	A	A	S	L	N	S	R	V
101	A	S	I	S	L	P	T	S	C	A	S	A	G	T	Q	C	L	I	S	G
121	W	G	N	T	K	S	S	G	T	S	Y	P	D	V	L	K	C	L	K	A
141	P	I	L	S	D	S	S	C	K	S	A	Y	P	G	Q	I	T	S	N	M
161	F	C	A	G	Y	L	E	G	G	K	D	S	C	Q	G	D	S	G	G	P
181	V	V	C	S	G	K	L	Q	G	I	V	S	W	G	S	G	C	A	Q	K
201	N	K	P	G	V	Y	T	K	V	C	N	Y	V	S	W	I	K	Q	T	I
221	A	S	N																	

AMINO ACID SEQUENCE OF TRYPSIN (EXCERPT FROM PDB)

Fig. 38 B

1	V	V	G	G	T	E	A	Q	R	N	S	W	P	S	Q	I	S	L	Q	Y
21	R	S	G	S	S	W	A	H	T	C	G	G	T	L	I	R	Q	N	W	V
41	M	T	A	A	H	C	V	D	R	E	L	T	F	R	V	V	V	G	E	H
61	N	L	N	Q	N	N	G	T	E	Q	Y	V	G	V	Q	K	I	V	V	
81	P	Y	W	N	T	D	D	V	A	A	G	Y	D	I	A	L	L	R	L	A
101	Q	S	V	T	L	N	S	Y	V	Q	L	G	V	L	P	R	A	G	T	I
121	L	A	N	S	P	C	Y	I	T	T	G	W	G	L	T	R	T	N	G	Q
141	L	A	Q	T	L	Q	Q	A	Y	L	P	T	V	D	Y	A	I	C	S	S
161	S	S	Y	W	G	S	T	V	K	N	S	M	V	C	A	G	G	D	G	V
181	R	S	G	C	Q	G	D	S	G	G	P	L	H	C	L	V	N	G	Q	Y
201	A	V	H	G	V	T	S	F	V	S	R	L	G	C	N	V	T	R	K	P
221	T	V	F	T	R	V	S	A	Y	I	S	W	I	N	N	V	I	A	S	N

AMINO ACID SEQUENCE OF ELASTASE (EXCERPT FROM PDB)

Fig. 39 A

```
Key site number 36 — 41 in Trypsin

41 42 43 44 45 46
 M  T  A  A  H  C   < target >
 V  S  A  A  H  C   < probe  >

d = 12.070038 [A]
r.m.s.d. = 0.061077 [A]
The number of atoms in a probe = 6
The number of atoms in PDB = 240
The number of combination = 1
Time = 1sec
```

RETRIEVED RESULTS OF HISTIDINE ACTIVE SITES

Fig. 39 B

```
Key site number 175 — 179 in Trypsin

186 187 188 189 190
  G   D   S   G   G   < target >
  G   D   S   G   G   < probe  >

d = 8.922721 [A]
r.m.s.d. = 0.092879 [A]
The number of atoms in a probe = 5
The number of atoms in PDB = 240
The number of combination = 1
Time = 1sec
```

RETRIEVED RESULTS OF SERINE ACTIVE SITES

Fig. 40

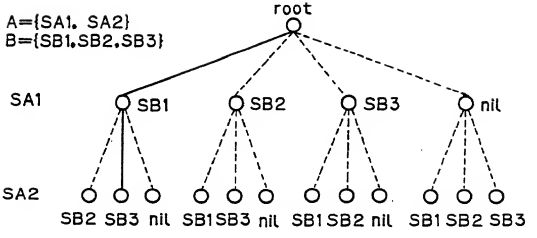

root
A={SA1, SA2}
B={SB1,SB2,SB3}
SA1
SB1
SB2
SB3
nil
SA2
SB2 SB3 nil
SB1 SB3 nil
SB1 SB2 nil
SB1 SB2 SB3

Fig. 41

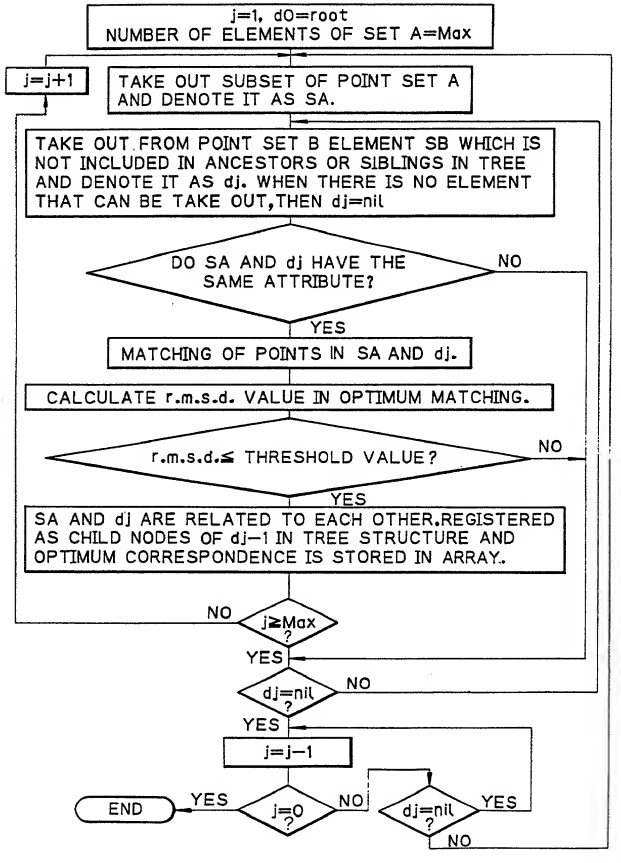

j=1, d0=root
NUMBER OF ELEMENTS OF SET A=Max
j=j+1
TAKE OUT SUBSET OF POINT SET A AND DENOTE IT AS SA.
TAKE OUT FROM POINT SET B ELEMENT SB WHICH IS NOT INCLUDED IN ANCESTORS OR SIBLINGS IN TREE AND DENOTE IT AS dj. WHEN THERE IS NO ELEMENT THAT CAN BE TAKE OUT,THEN dj=nil
DO SA AND dj HAVE THE SAME ATTRIBUTE?
NO
YES
MATCHING OF POINTS IN SA AND dj.
CALCULATE r.m.s.d. VALUE IN OPTIMUM MATCHING.
r.m.s.d.≦ THRESHOLD VALUE?
NO
YES
SA AND dj ARE RELATED TO EACH OTHER.REGISTERED AS CHILD NODES OF dj−1 IN TREE STRUCTURE AND OPTIMUM CORRESPONDENCE IS STORED IN ARRAY.
NO
j≧Max ?
YES
dj=nil ?
NO
YES
j=j−1
END
YES
j=0 ?
NO
dj=nil ?
YES
NO

Fig. 42

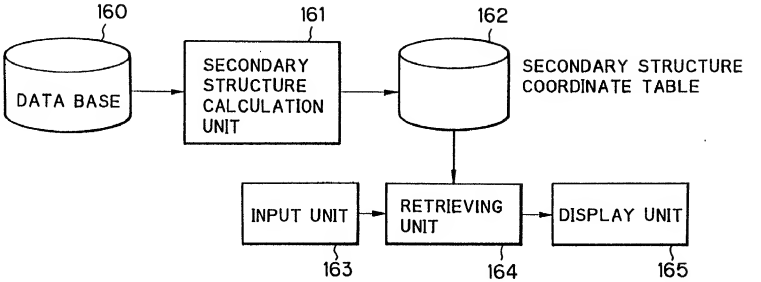

160
161
162
DATA BASE
SECONDARY STRUCTURE CALCULATION UNIT
SECONDARY STRUCTURE COORDINATE TABLE
INPUT UNIT
RETRIEVING UNIT
DISPLAY UNIT
163
164
165

Fig. 43

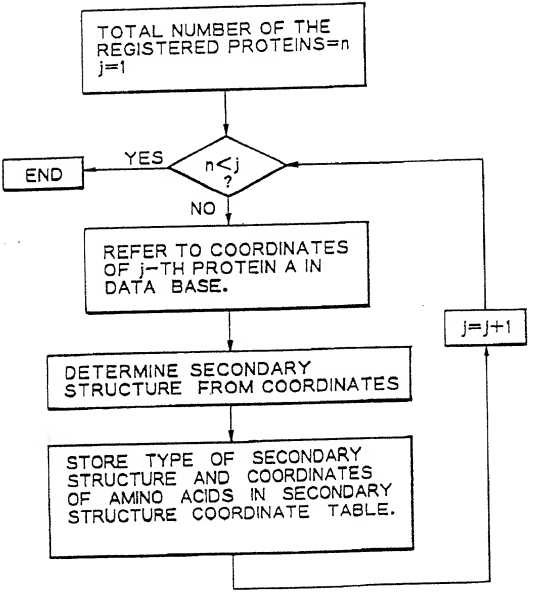
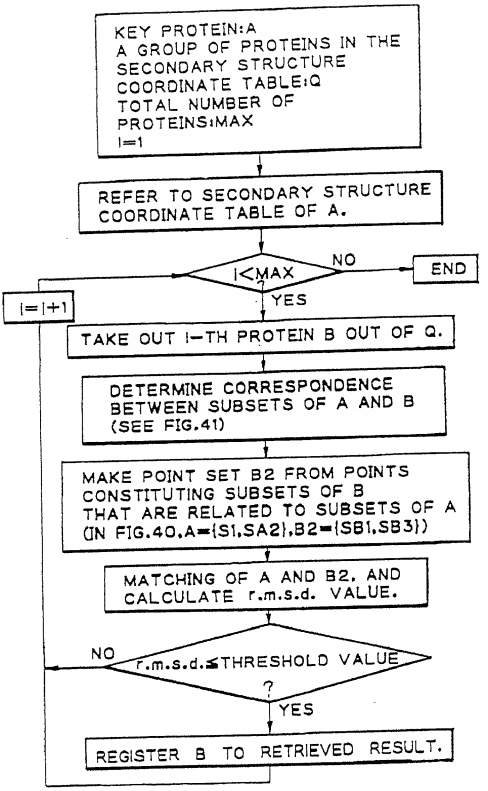

TOTAL NUMBER OF THE REGISTERED PROTEINS=n
j=1
n<j ?
YES
END
NO
REFER TO COORDINATES OF j-TH PROTEIN A IN DATA BASE.
DETERMINE SECONDARY STRUCTURE FROM COORDINATES
STORE TYPE OF SECONDARY STRUCTURE AND COORDINATES OF AMINO ACIDS IN SECONDARY STRUCTURE COORDINATE TABLE.
j=j+1

Fig. 44

162

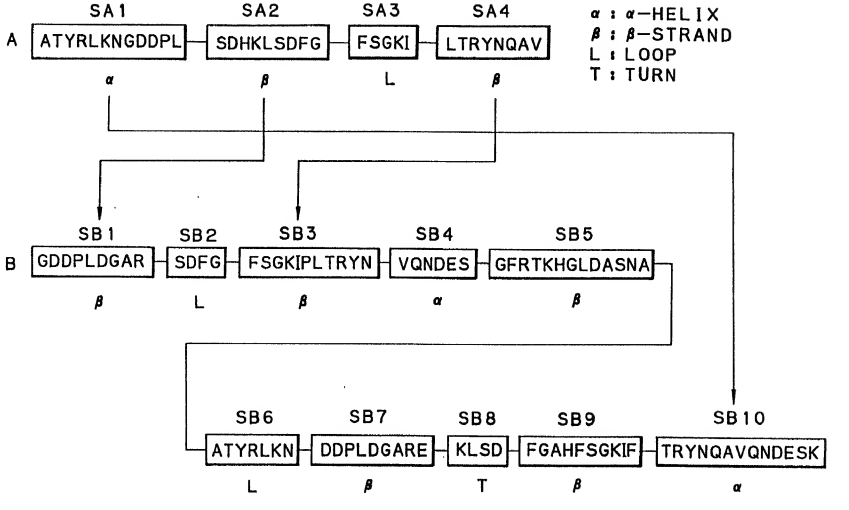
SUBSET	COORDINATES	TYPE
S1	{X1, X2, X3, X4, ····· Xa}	α — HELIX
S2	{Xa+1, Xa+2, ····· Xb}	α — HELIX
S3	{Xb+1, Xb+2, ····· Xc}	β — SHEET
S4	{Xc+1, Xc+2, ····· Xd}	β — SHEET
	⋮	⋮
Sn	{Xl+1, Xl+2, ····· Xm}	3 — TURN

Fig. 45


KEY PROTEIN:A
A GROUP OF PROTEINS IN THE SECONDARY STRUCTURE COORDINATE TABLE:Q
TOTAL NUMBER OF PROTEINS:MAX
I=1
REFER TO SECONDARY STRUCTURE COORDINATE TABLE OF A.
I<MAX ?
NO
END
YES
I=I+1
TAKE OUT I-TH PROTEIN B OUT OF Q.
DETERMINE CORRESPONDENCE BETWEEN SUBSETS OF A AND B (SEE FIG.41)
MAKE POINT SET B2 FROM POINTS CONSTITUTING SUBSETS OF B THAT ARE RELATED TO SUBSETS OF A (IN FIG.40, A={S1,SA2}, B2={SB1,SB3})
MATCHING OF A AND B2, AND CALCULATE r.m.s.d. VALUE.
NO
r.m.s.d.≤THRESHOLD VALUE ?
YES
REGISTER B TO RETRIEVED RESULT.

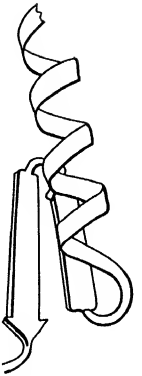
099100051.072301

Fig. 46



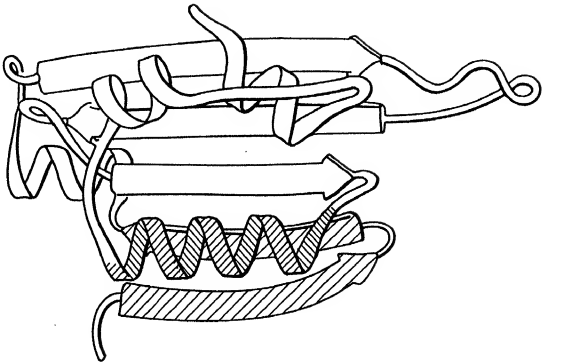
099100514.072301

Fig. 47 A



KEY PROTEIN A

Fig. 47 B



PROTEIN B HAVING A SIMILAR STRUCTURE